

Thermodynamics of the six-vertex model in an L-shaped domain

Filippo Colomo and Andrei G. Pronko

ABSTRACT. We consider the six-vertex model in an L-shaped domain of the square lattice, with domain wall boundary conditions. For free-fermion vertex weights the partition function can be expressed in terms of some Hankel determinant, or equivalently as a Coulomb gas with discrete measure and a non-polynomial potential with two hard walls. We use Coulomb gas methods to study the partition function in the thermodynamic limit. We obtain the free energy of the six-vertex model as a function of the parameters describing the geometry of the scaled L-shaped domain. Under variations of these parameters the system undergoes a third-order phase transition. The result can also be considered in the context of dimer models, for the perfect matchings of the Aztec diamond graph with a cut-off corner.

1. Introduction

It is commonly known that macroscopic quantities (such as the free energy) of dimer coverings and random tilings on regular lattices may depend on boundary conditions [1, 2]. This feature is related to phase separation and the emergence of a limit shape [3–7], see [8] for a review. The same phenomena can be observed in the six-vertex model [9–11]. In relation with these phenomena, an interesting question concerns the stability of the observed bulk properties against various deformations of the shape of the considered finite region, while preserving the particular boundary conditions (e.g., the staircase shape for the boundary of the Aztec diamond in the case of domino tilings).

In this paper, we address the problem by studying the thermodynamics of the six-vertex model on a particular domain of the square lattice, that we call an L-shaped domain. The model is closely related to the domino tilings of the Aztec diamond with a cut-off corner [12]. We evaluate the exact analytic expression for the free energy per site of the six-vertex model at its free fermion point, as a function of the parameters describing the geometry of the scaled L-shaped domain. Under variation of these parameters, when the boundary interferes with the phase separation curve, the system undergoes a third-order phase transition.

Our result on the free energy of the six-vertex model in an L-shaped domain is based on some discrete Coulomb gas representation. Discrete Coulomb gases already appeared in applications to tiling and dimer models [13, 14], and to the six-vertex model [15, 16], and have been studied intensively in the context of discrete orthogonal polynomials, see [17] and references therein. As a side result, we find

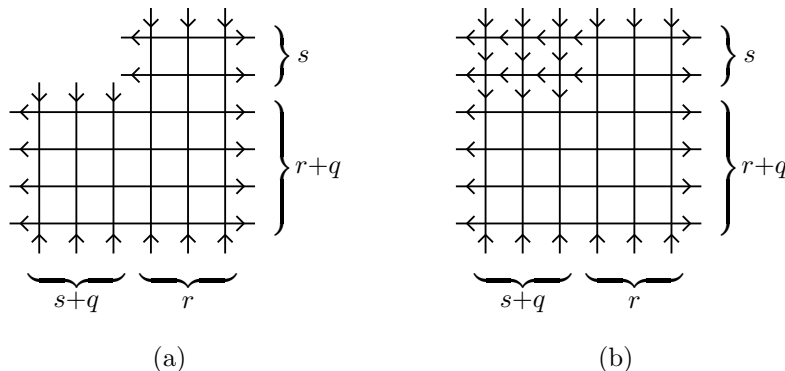


FIGURE 1. (a) The L-shaped domain with domain wall boundary conditions; (b) The corresponding arrow configuration on the $N \times N$ lattice.

that the soft-edge/hard-edge transition may induce (or not) a third-order phase transition in the free energy, according to the finite (or diverging) slope of the discrete Coulomb gas potential in the vicinity of the hard wall.

1.1. The model. The states of the six-vertex model are configurations of arrows on the edges of a square lattice, satisfying the *ice rule*: at each vertex the numbers of incoming and outgoing arrows are equal. The ice rule selects $\binom{4}{2} = 6$ possible vertex configurations, hence the name of the model.

The L-shaped domain can be defined as a square domain with a rectangular portion removed from one of the corners, see Fig. 1. Specifically, the square domain is the finite square lattice obtained from the intersection of N horizontal and N vertical lines, the so-called $N \times N$ lattice. The L-shaped domain is obtained by removing from the top left corner a rectangular portion of the lattice, of size $s \times (s+q)$, $s, q \in \mathbb{N}_0$, $s+q < N$. The L-shaped domain is shown in Fig. 1a, where $r = N - s - q$.

We consider a specific instance of fixed boundary conditions, the domain wall boundary conditions. In the standard arrows picture of the six-vertex model (see, e.g., [18]) this conditions means that all horizontal (respectively, vertical) arrows on external edges are outgoing (incoming). For $s = 0$, one has the usual $N \times N$ lattice with domain wall boundary conditions, introduced in [19].

The Boltzmann weights of the six-vertex model, w_i , $i = 1, \dots, 6$, enumerated in the standard order, see Fig. 2, second row, are chosen to satisfy

$$w_1 w_2 + w_3 w_4 = w_5 w_6, \quad (1.1)$$

that is called the free-fermion condition.

We denote the partition function of the six-vertex model in an L-shaped domain with domain wall boundary conditions by $Z_{r,s,q}$; for the special case $s = 0$, we use the standard notation Z_N . For generic Boltzmann weights, the partition function Z_N has been evaluated in determinantal form in [20, 21]. For weights satisfying the condition (1.1), Z_N has an extremely concise form:

$$Z_N = w_5^{\frac{N(N-1)}{2}} w_6^{\frac{N(N+1)}{2}}.$$

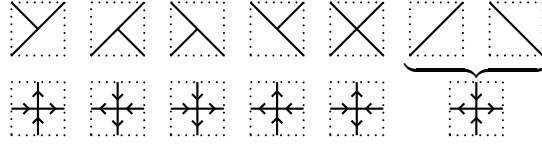


FIGURE 2. Small patches of dominoes (first row) and vertex configurations of the six-vertex model (second row).

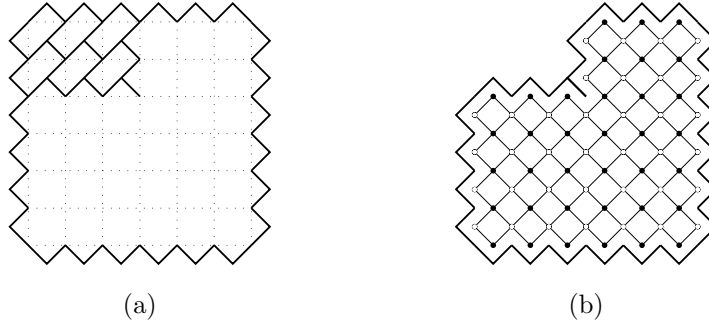


FIGURE 3. (a) Aztec diamond with a frozen region of NE-SW dominoes; (b) Aztec diamond with a cut-off corner and the corresponding modified Aztec diamond graph.

For a proof, see [2] in the context of enumerative combinatorics, or [22] in the context of integrable models.

For generic s , the partition function $Z_{r,s,q}$ can be expressed in terms of certain non local correlation function of the model on the $N \times N$ lattice. This function is called the emptiness formation probability, and denoted here as $F_{r,s,q}$. It describes the probability that all vertices in the top left $s \times (s+q)$ rectangle of the $N \times N$ lattice are of type 2. The corresponding configuration is shown in Fig. 1b. Comparing the two pictures of Fig. 1, it is clear that

$$Z_{r,s,q} = \frac{Z_N}{w_2^{s(s+q)}} F_{r,s,q}. \quad (1.2)$$

Thus the partition function of the six-vertex model on the L-shaped domain is essentially given by the emptiness formation probability of the model on the $N \times N$ lattice.

The above construction can be translated in the language of dimer models, using the well known correspondence between the six-vertex model with domain wall boundary conditions and the domino tilings of the Aztec diamond [2].

Recall that *domino tilings* are coverings of a square lattice with rectangles of size 1×2 and 2×1 . Domino tilings can be formulated in terms of the six-vertex model on a square lattice by mapping elementary patches of domino tilings to arrow configurations as shown in Fig. 2. The case of the six-vertex model on an $N \times N$ square lattice with domain wall boundary conditions corresponds to the Aztec diamond of order N , see Fig. 3.

The plain enumeration of domino tilings of the Aztec diamond corresponds to the choice of the Boltzmann weights of the six-vertex model $w_1 = \dots = w_5 =$

1 and $w_6 = 2$. In counting of domino tilings, one may also consider weighted enumerations, with a *bias* parameter $\alpha \in [0, 1]$, describing the asymmetry between the two possible orientations of dominoes. Following paper [3], we assign to the NE-SW dominoes the weight $\sqrt{2(1-\alpha)}$, and to the NW-SE the weight $\sqrt{2\alpha}$. To establish the correspondence with the six-vertex model, it is useful to parameterize the Boltzmann weights as follows:

$$w_1 = w_2 = \sqrt{\rho(1-\alpha)}, \quad w_3 = w_4 = \sqrt{\rho\alpha}, \quad w_5 = 1, \quad w_6 = \rho. \quad (1.3)$$

Then the weighted enumeration of dominoes corresponds to setting $\rho = 2$.

In the same spirit, the six-vertex model in an L-shaped domain can be related to the domino tilings of some suitable modification of the Aztec diamond. Using the correspondence of vertices of type 2 with a particular patch of domino tilings, see Fig. 2, it is clear that emptiness formation probability is equivalent to the probability of observing the domino configuration shown in Fig. 3a. Removing the frozen region of NE-SW dominoes from the Aztec diamond, one obtain the Aztec diamond with a cut-off corner shown in Fig. 3b. The same pictures also shows the resulting graph for dimer coverings (modified Aztec diamond graph).

1.2. Discrete Coulomb gas representation. As discussed above, the partition function of the six-vertex model on an L-shaped domain is essentially given by the emptiness formation probability. For generic values of Boltzmann weights, the emptiness formation probability has been computed in the form of a multiple integral in [23]. Under the free-fermion condition this representation reduces to an Hankel determinant [24].

PROPOSITION 1.1. *The emptiness formation probability of the six-vertex model on the $N \times N$ lattice with domain wall boundary condition, with the Boltzmann weights (1.3) admits the following representation:*

$$F_{r,s,q} = \frac{(q!)^s}{\prod_{k=0}^{s-1} (q+k)!k!} \frac{(1-\alpha)^{s(s+q)}}{\alpha^{s(s-1)/2}} \det_{1 \leq j,k \leq s} \left[\sum_{m=0}^{r-1} m^{j+k-2} \binom{m+q}{m} \alpha^m \right]. \quad (1.4)$$

For a proof, see [23, 24].

REMARK 1.2. *An alternative representation, equivalent to (1.4), reads*

$$F_{r,s,q} = \frac{(1-\alpha)^{s(s+q)}}{\alpha^{s(s-1)/2}} I_{r,s,q}, \quad (1.5)$$

where

$$I_{r,s,q} = \frac{1}{s!} \prod_{j=0}^{s-1} \frac{q!}{j!(j+q)!} \sum_{m_1=0}^{r-1} \cdots \sum_{m_s=0}^{r-1} \prod_{1 \leq j < k \leq s} (m_j - m_k)^2 \prod_{j=1}^s \binom{q+m_j}{q} \alpha^{m_j}. \quad (1.6)$$

The last representation is that of a discrete Coulomb gas confined within a finite interval. We recognize in the discrete weight

$$w_q^\alpha(m) = \binom{q+m}{q} \alpha^m, \quad m \in \mathbb{N}_0,$$

that of Meixner polynomials, but we emphasize the condition $m < r$ for the charge coordinates in the Coulomb gas representation (1.6). This condition plays a major role in what follows. We shall refer to this condition as an *hard wall* for the charges.

The discrete Coulomb gas formula (1.6) appeared previously in the context of some random growth model studied by Johansson [13]. The connections of (1.6) with the circular unitary ensemble, with a multivariate generalization of hypergeometric function, and with the τ -function of the sixth Painlevé equation were enlightened by Forrester and Witte [25].

1.3. Thermodynamics: the main result. We are interested in the behaviour of the model in the thermodynamic limit, i.e., in the limit of the large lattice sizes, $r, s, q \rightarrow \infty$. This is described by the free energy per site, defined as follows. Let $s + q = [xN]$, $s = [yN]$, with $x, y \in [0, 1]$ fixed, $x \geq y$. We set

$$f(x, y) := - \lim_{N \rightarrow \infty} \frac{1}{N^2 - s(s + q)} \log Z_{N-s-q, s, q}. \quad (1.7)$$

Equivalently, we may study the emptiness formation probability $F_{r, s, q}$. We set

$$\varphi(x, y) := - \lim_{N \rightarrow \infty} \frac{1}{N^2} \log F_{N-s-q, s, q}. \quad (1.8)$$

REMARK 1.3. *The above limits exist; in particular, $\varphi(x, y)$ is non-negative all over its domain of definition, $x, y \in [0, 1]$, $x \geq y$.*

This is part of a stronger statement proven in [13], see Theorem 1.1 therein.

From relation (1.2) it immediately follows that the free energy per site of the six-vertex model on the L-shaped domain is completely determined by the knowledge of the function $\varphi(x, y)$. Indeed, one has

$$(1 - xy)f(x, y) = f(0, 0) + xy \log w_2 + \varphi(x, y), \quad f(0, 0) = -\log \sqrt{\rho}.$$

Similarly, the function $\varphi(x, y)$ determines completely the free energy of the domino tilings (up to trivial modifications, which can be easily inferred from (1.2)).

The main result of the present paper concerns the explicit form of the function $\varphi(x, y)$.

Let us consider the unit square, parameterized by $x, y \in [0, 1]$. It follows from (1.8) that the domain of definition of the function $\varphi(x, y)$ is the triangular region delimited by the three lines $y = 0$, $y = x$, and $y = 1 - x$. Let us consider the ellipse defined by the equation

$$\frac{(1 - x - y)^2}{\alpha} + \frac{(x - y)^2}{1 - \alpha} = 1. \quad (1.9)$$

We denote by \mathcal{D}_I and \mathcal{D}_{II} the two regions of the triangle that are inside, respectively outside, the ellipse, see Fig. 4. The arc of the ellipse (1.9) lying in the triangle, and separating these two regions, is described by the equation:

$$\sqrt{y} = \sqrt{(1 - x)(1 - \alpha)} - \sqrt{x\alpha}, \quad 0 \leq y \leq x \leq 1 - \alpha. \quad (1.10)$$

We denote this arc by \mathcal{A} .

THEOREM 1.4. *The function $\varphi(x, y)$, $(x, y) \in [0, 1] \times [0, 1]$, is*

$$\varphi(x, y) = 0, \quad (x, y) \in \mathcal{D}_I,$$

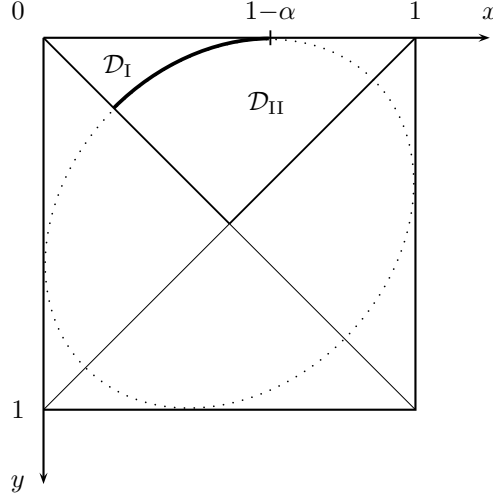


FIGURE 4. The Arctic ellipse (dotted), the arc \mathcal{A} (bold), and the domains \mathcal{D}_I and \mathcal{D}_{II} .

and

$$\begin{aligned} \varphi(x, y) = & xy \log \frac{h}{\eta} - \frac{(1-x-y)^2}{2} \log \frac{1-h}{1-\eta} - \frac{1}{2} \log \frac{1+h}{1+\eta} \\ & + (1-x)y \log \frac{y+(1-x)h}{y+(1-x)\eta} + x(1-y) \log \frac{x+(1-y)h}{x+(1-y)\eta} \\ & - (1-x)x \log \frac{x+(1-x)h}{x+(1-x)\eta} - (1-y)y \log \frac{y+(1-y)h}{y+(1-y)\eta} \\ & - \frac{(x-y)^2}{2} \log \frac{x+y+(2-x-y)h}{x+y+(2-x-y)\eta}, \quad (x, y) \in \mathcal{D}_{II}, \end{aligned} \quad (1.11)$$

where

$$h = h(x, y) := \sqrt{\frac{xy}{(1-x)(1-y)}} \quad (1.12)$$

and $\eta = \eta(x, y; \alpha)$ is such that

$$\eta \in [0, 1], \quad \alpha \frac{(1+\eta)^2 (x+(1-x)\eta)(y+(1-y)\eta)}{(1-\eta)^2 (y+(1-x)\eta)(x+(1-y)\eta)} = 1. \quad (1.13)$$

The fact that $\varphi(x, y)$ vanishes in the domain \mathcal{D}_I was already proven in [13]. A few comments are in order.

REMARK 1.5. *The function $\varphi(x, y)$ is a continuous function of its variables. In particular it vanishes on the curve \mathcal{A} :*

$$\varphi(x, y)|_{(x,y) \in \mathcal{A}} = 0. \quad (1.14)$$

This property can be verified as follows. Let $(x, y) \in \mathcal{A}$, then $\alpha = \alpha_c$, where

$$\alpha_c = \alpha_c(x, y) := \left(\sqrt{(1-x)(1-y)} - \sqrt{xy} \right)^2. \quad (1.15)$$

Setting $\alpha = \alpha_c$ in (1.13), one can directly check that the root of equation (1.13) is $\eta = h$, with h given by (1.12). In other words,

$$\eta(x, y; \alpha)|_{(x, y) \in \mathcal{A}} = \eta(x, y; \alpha_c) = h(x, y), \quad (1.16)$$

and (1.14) follows from (1.11).

The function $\varphi(x, y) = \varphi(x, y; \alpha)$ may also be viewed as a function of α at fixed x and y .

REMARK 1.6. *The function $\varphi(x, y)$ satisfies*

$$\varphi(x, y; \alpha_c) = 0.$$

This property follows from the second equality in (1.16). In fact, stronger properties hold.

PROPOSITION 1.7. *Given (x, y) , for values of α in the vicinity of $\alpha_c = \alpha_c(x, y)$, we have*

$$\varphi(x, y) = \begin{cases} 0 & \alpha < \alpha_c \\ C(\alpha - \alpha_c)^3 + O((\alpha - \alpha_c)^4) & \alpha > \alpha_c, \end{cases}$$

where $C = C(x, y) > 0$.

PROPOSITION 1.8. *Given a point $(x, y) \in \mathcal{D}_{\text{II}}$, at a distance t from the arc of ellipse \mathcal{A} , one has*

$$\varphi(x, y) \propto t^3, \quad t \rightarrow 0^+.$$

where the proportionality constant is positive.

These properties follow from the explicit form of function $\varphi(x, y)$ given in Theorem 1.4.

1.4. Discussion. The ellipse (1.9) is easily recognized as the Arctic ellipse of the domino tilings of the original, non-deformed Aztec diamond [3], or, equivalently, of the six-vertex model with domain wall boundary condition, at its free-fermion point, on the original square domain.

The function $\varphi(x, y)$ vanishes at $\alpha = \alpha_c$, together with its first and second derivatives with respect to α , and has non-vanishing (and finite) third-order derivative. This property obviously holds also for the free energy $f(x, y)$, see (1.7). Recalling that the parameter α is directly related to the Boltzmann weights, we thus observe the occurrence of a third-order phase transition at $\alpha = \alpha_c$.

An alternative point of view is that of considering α fixed, while varying the shape of the L-shaped domain, and thus the coordinates x and y . The function $\varphi(x, y)$, studied as a function of the parameter t , again undergoes a third-order phase transition.

This result leads to an alternative interpretation of the Arctic ellipse, as a critical curve in the space of the parameters of the model, a point of view already advocated in [10]. The result extends that reported in [12], where the analysis was restricted to the line $x = y$, and only the critical point $(x, y) = (\frac{1-\sqrt{\alpha}}{2}, \frac{1-\sqrt{\alpha}}{2})$ could be observed.

The vanishing of the function $\varphi(x, y)$ in domain \mathcal{D}_{I} is somewhat obvious, since it is known that in \mathcal{D}_{I} , the emptiness formation probability is equal to 1 up to exponentially small corrections, which are related to the so-called upper tail rate function [13]. The function $\varphi(x, y)$ in domain \mathcal{D}_{II} defines instead the so-called lower tail rate function, whose explicit form is given by (1.11).

In the derivation of Theorem (1.4) we rely heavily on representation (1.6), that is a Coulomb gas on a discrete lattice, confined by a non-polynomial potential with two hard walls. Thus, a by-product of our results is that the hard wall in the discrete Coulomb gas can induce a soft-edge/hard-edge transition with a third-order discontinuity in the corresponding free energy. This is a well-known phenomenon in the case of a continuous Coulomb gas, see [26] and references therein. However, in the discrete case, phase transitions are rather of the Douglas-Kazakov type [27], while the soft-edge/hard-edge transition is usually *transparent*, in that it does not come with any discontinuity in the free energy (see [17] for numerous examples).

The peculiar behaviour observed here can be ascribed to the form of the potential in the vicinity of the hard walls. In previously studied cases, in the vicinity of a hard wall situated, say, at x_0 , the derivative of the potential has a divergent behaviour $V'(x) \simeq \log|x - x_0|$, implying a smooth matching of the potential with the hard wall. In the present model, instead, such smooth matching occurs only in correspondence of the left hard wall. As a result, among the two possible soft-edge/hard-edge transitions in the present model only one is transparent, that is related to the left hard wall. The other one, related to the right hard wall, comes indeed with a third-order phase transition in the corresponding free energy.

Acknowledgments. We thank A. Kuijlaars and P. Zinn-Justin for useful discussions. This work is partially supported by the IRSES grant of EC-FP7 Marie Curie Action “Quantum Integrability, Conformal Field Theory and Topological Quantum Computation”. A.G.P. acknowledges partial support from the Russian Foundation for Basic Research, grant 13-01-00336, and from INFN, Sezione di Firenze.

2. The equilibrium measure

We first outline some general aspects of the asymptotic behaviour of the discrete Coulomb gas, along the lines developed in [17], see also [13, 16], and next switch to the discussion of explicit solutions for the resolvent of the equilibrium measure. The end-point equations and the free energy are discussed in the next section.

2.1. The discrete Coulomb gas. In order to evaluate the limit (1.8) and prove Theorem 1.4, we study the asymptotic behaviour of the discrete Coulomb gas (1.6) in the scaling limit, that is $r, s, q \rightarrow \infty$, while preserving the aspect ratio of the L-shaped domain. Let $r = [Rs]$, $R \geq 1$ fixed and $q = [Qs]$, $Q \geq 0$ fixed. We set

$$\Phi(R, Q) := \lim_{s \rightarrow \infty} \frac{1}{s^2} \log I_{r, s, q}. \quad (2.1)$$

It was shown in [13], see Theorem 2.2 therein, that the above limit exists. The function $\Phi(R, Q)$ is related to the function $\varphi(x, y)$ introduced in (1.8) by

$$(1 - x)^2 \varphi(x, y) = -xy \log(1 - \alpha) + y^2 \log \sqrt{\alpha} - y^2 \Phi\left(\frac{1 - x}{y}, \frac{x - y}{y}\right), \quad (2.2)$$

see (1.5). To obtain relation (2.2) one has to take into account the relations

$$R = \frac{1 - x}{y}, \quad Q = \frac{x - y}{y}, \quad (2.3)$$

which hold in the limit $s \rightarrow \infty$.

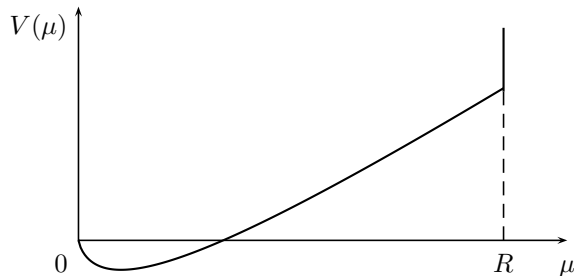


FIGURE 5. The potential $V(\mu)$ with two hard walls, at $\mu = 0$ and $\mu = R$.

In [13], the focus was on the distribution of the position of the rightmost charge, $\max_{1 \leq j \leq s} m_j$, that was shown to follow, in the limit $s \rightarrow \infty$, the Tracy-Widom distribution [28]. In [13], the explicit form of the upper tail rate function was evaluated. Our aim here is to evaluate the explicit form of function $\Phi(R, Q)$, and hence, the lower tail rate function $\varphi(x, y)$.

As discussed in [13] to evaluate the continuous limit of the discrete Coulomb gas (1.6) we may just as well consider the corresponding continuous Coulomb gas, in a continuous potential derived from the original discrete measure. Set

$$V_s(\mu) := -\frac{1}{s} \log \left[\binom{q + s\mu}{q} \alpha^{s\mu} \right], \quad \mu \geq 0.$$

Using Stirling formula, we have

$$V(\mu) := \lim_{s \rightarrow \infty} V_s(\mu) = -\mu \log \alpha + \mu \log \mu - (Q + \mu) \log(Q + \mu) + Q \log Q, \quad (2.4)$$

uniformly on compact subsets of $[0, \infty)$. Note that the confining potential $V(\mu)$ is complemented by two hard walls, at $\mu = 0$ and $\mu = R$, see Fig. 5. We emphasize the difference in the behaviour of the potential in the vicinity of the two hard walls: the derivative of the potential diverges logarithmically at $\mu = 0$, while it stays *finite* at $\mu = R$. As a result, the potential joins up the hard wall smoothly on the left, but with a corner on the right.

To proceed, we rely on some known results from weighted potential theory in the case of a discrete Coulomb gas. This has been discussed both in the physical [27] and mathematical literature [29, 30]; see [17] for a general mathematical treatment. In a nutshell, under mild assumptions on the potential that are clearly satisfied in our problem, see [13], the asymptotic behaviour of the discrete Coulomb gas is governed by a constrained equilibrium problem analogous to that for the continuous Coulomb gas, with the additional constraint that the particle density is always less than or equal to 1, as a consequence of discreteness.

More specifically, the problem reduces to the evaluation of the equilibrium measure, $\rho_0(\mu)d\mu$, that is the unique minimizer of the functional

$$S[\rho] = - \int \int_{\mu \neq \nu} \log |\mu - \nu| \rho(\mu) \rho(\nu) d\mu d\nu + \int V(\mu) \rho(\mu) d\mu,$$

subject to the constraint (normalization condition)

$$\int \rho(\mu) d\mu = 1, \quad (2.5)$$

and to the additional constraint

$$\rho(\mu) \leq 1, \quad (2.6)$$

that is imposed by the discreteness of the measure of the Coulomb gas. In the above formulas, integration is understood over the interval $[0, R]$.

It is known that, due to the presence of the last constraint, the equilibrium measure partitions the interval $[0, R]$ into three sets, I_- , I_0 , I_+ , according to $\rho_0(\mu) = 0$, $0 < \rho_0(\mu) < 1$, or $\rho_0(\mu) = 1$, respectively. Each set is the union of a finite number of intervals, named *voids*, *bands*, or *saturated regions*, respectively. Voids and saturated regions are collectively named *gaps*.

2.2. The resolvent. The most convenient way to evaluate the equilibrium measure is by introducing its *resolvent*

$$W(z) = \int_0^R \frac{\rho_0(\mu)}{z - \mu} d\mu, \quad z \notin [0, R],$$

that has the following properties:

- (1) $W(z)$ is analytic in $\mathbb{C} \setminus [0, R]$.
- (2) For large z , the resolvent has the asymptotic behaviour

$$W(z) = \frac{1}{z} + \frac{E}{z^2} + O(z^{-3}), \quad |z| \rightarrow \infty, \quad (2.7)$$

where the leading coefficient is determined by the normalization condition (2.5), and E is the first moment of the equilibrium measure:

$$E = \int_0^R \mu \rho_0(\mu) d\mu. \quad (2.8)$$

- (3) The particle density can be found as the discontinuity of the resolvent across its cuts, $I_0 \cup I_+$,

$$\rho_0(\mu) = -\frac{1}{2\pi i} [W(\mu + i0) - W(\mu - i0)], \quad \mu \in [0, R]. \quad (2.9)$$

In particular, it follows that

$$W(\mu + i0) - W(\mu - i0) = \begin{cases} 0 & \mu \in I_- \\ -2\pi i & \mu \in I_+. \end{cases}$$

- (4) The resolvent satisfies

$$W(\mu + i0) + W(\mu - i0) = V'(\mu), \quad \mu \in I_0. \quad (2.10)$$

The above properties, together with some ‘ansatz’ on the structure of the partition of the interval $[0, R]$ into voids, bands, and saturation regions, are sufficient to completely determine the resolvent associated to the considered discrete Coulomb gas.

Before discussing the issues related to the discrete nature of the Coulomb gas, and to the related phenomenon of emergence of saturated regions, it is useful to recall the procedure for the determination of the resolvent in a case where the discreteness and hard walls play no role, such as that of a single band $[a, b]$, with two voids, $[0, a]$ and $[b, R]$. The equation for the resolvent is a singular integral equation of the form (2.10), where $I_0 = [a, b]$. The end-points of the band are determined by the condition that the solution of (2.10) should match at leading order the large z behaviour $W(z) \sim 1/z$, dictated by (2.7).

However, if the density found by solving (2.10) does not satisfy the condition (2.6), or the end-points a and b are not internal to the interval $[0, R]$, then one has to introduce some suitable ansatz about the structure of the partition of the interval $[0, R]$ into voids, bands, and saturated regions, with I_+ consisting of at least one interval.

The existence of at least one saturated region (that is, $I_+ \neq \emptyset$) implies the following form of the resolvent:

$$W(z) = \int_{I_+} \frac{1}{z - \mu} d\mu + H(z), \quad H(z) := \int_{I_0} \frac{\rho_0(\mu)}{z - \mu} d\mu.$$

Here the function $H(z)$ can be found from the equation

$$H(\mu + i0) + H(\mu - i0) = U(\mu), \quad \mu \in I_0, \quad (2.11)$$

where

$$U(z) = -2 \int_{I_+} \frac{1}{z - \mu} d\mu + V'(z).$$

Assuming that I_0 consists of just a single interval (the so-called one-band ansatz, [17]), with the end-points a and b satisfying $0 < a < b < R$, the solution of (2.11) such that $H(z) = O(1/z)$ as $z \rightarrow \infty$, is

$$H(z) = \frac{\sqrt{(z-a)(z-b)}}{2\pi} \int_a^b \frac{U(\mu)}{(z-\mu)\sqrt{(\mu-a)(b-\mu)}} d\mu. \quad (2.12)$$

The end-points a and b can be found by solving the equations on the coefficients of order z^0 and z^{-1} terms of the large z expansion of (2.12), to match the the large z behaviour (2.7) at leading order. Finally, the explicit form of the particle density is derived by using (2.9), and its consistency against condition (2.6) may be verified.

In formulating some ansatz on the structure of the partition of the interval $[0, R]$ into voids, bands, and saturation regions, we take into account the form of the potential (2.4), together with the presence of two hard walls (that is infinite potential barriers) at $\mu = 0, R$. It is clear that the band-gap structure may be affected by varying the parameters $R \in [1, \infty)$ and $Q \in [0, \infty)$. The whole solution is obtained by investigating how the band-gap structure is modified as these parameters are varied.

Concerning the shape of the potential (2.4), it is clear that $V(0) = 0$ and, as $\mu \rightarrow \infty$, the function $V(\mu)$ is asymptotically linear, $V(\mu) \sim -(\log \alpha)\mu$, with a positive slope (recall that $0 < \alpha < 1$). The derivative of the potential reads

$$V'(\mu) = -\log \alpha + \log \mu - \log(\mu + Q).$$

Thus the function $V(\mu)$ has a single minimum in the interval $[0, R]$, at $\mu = \mu_0$, where

$$\mu_0 = \frac{Q\alpha}{1-\alpha}.$$

A nonvanishing density (with bands and possibly saturated regions) is thus expected in the vicinity of μ_0 .

Concerning the presence of hard walls, it is known that these affect the band-gap structure in that they admit only gaps in their vicinity [17]. Thus, in our case, the intervals adjacent to the points $\mu = 0$ and $\mu = R$ can only be voids or saturated regions. It is clear, basing on the physical picture of the Coulomb gas, that the

presence of a void or a saturated region is related to the distance of the minimum μ_0 from the hard walls at $\mu = 0, R$.

A detailed analysis provided below shows that four possible band-gap structures, or *scenarios* may occur. We refer to them as Regime IA, Regime IB, Regime IIA, and Regime IIB for definiteness. In all these scenarios, there is just one band, that we denote $[a, b]$, between two gaps, $[0, a]$ and $[b, R]$. The four scenarios differ one another in the nature (void or saturated region) of the two gaps. Namely, in Regime IA the first gap, $[0, a]$, is a saturated region, while the second gap, $[b, R]$, is a void; in Regime IB both gaps are voids; in Regime IIA both gaps are saturated regions; finally, in Regime IIB the first gap is a saturated region while the second is a void. The four regimes correspond to four different regions in the domain $Q \geq 0$, $R \geq 1$.

2.3. Regime I. Regime IA and Regime IB are closely related. Actually, they can be viewed as two different scenarios for one same regime, Regime I, characterized by a void on the right gap $[b, R]$. The equilibrium measures for Regime IA and Regime IB were first derived in [13].

PROPOSITION 2.1. *The resolvent of the equilibrium measure for the continuum limit of the discrete Coulomb gas (1.6) in Regime IA is*

$$W(z) = -\log \sqrt{\alpha} - \log \frac{\sqrt{a(z-b)} + \sqrt{b(z-a)}}{\sqrt{(b-a)z}} - \log \frac{\sqrt{(a+Q)(z-b)} + \sqrt{(b+Q)(z-a)}}{\sqrt{(b-a)z}}. \quad (2.13)$$

where the parameters a and b , the two band end-points, are

$$a = \frac{(1 - \sqrt{\alpha(1+Q)})^2}{1 - \alpha}, \quad b = \frac{(1 + \sqrt{\alpha(1+Q)})^2}{1 - \alpha}. \quad (2.14)$$

In Regime IA, there is one band $[a, b]$, with a saturated region $[0, a]$, and a void $[b, R]$. The saturated region in the interval $[0, a]$ implies that the resolvent has the form

$$W(z) = \log \frac{z}{z-a} + H(z), \quad H(z) := \int_a^b \frac{\rho_0(\mu)}{z-\mu} d\mu.$$

The function $W_{\text{IA}}(z)$ should solve equation (2.11) with $U(z) = V'(z)$ and $I_0 = [a, b]$, where the function $H(z)$ should solve (2.11), with function $U(\mu)$ chosen as

$$U(\mu) = -\log \alpha + \log \frac{(\mu-a)^2}{\mu(\mu+Q)}.$$

Evaluating the integral in (2.12) with this function and coming back to function $W(z)$, we obtain expression (2.13).

The density in the interval $[a, b]$ is given by the expression

$$\rho_0(\mu) = \frac{1}{\pi} \arctan \sqrt{\frac{a(b-\mu)}{b(\mu-a)}} + \frac{1}{\pi} \arctan \sqrt{\frac{(a+Q)(b-\mu)}{(b+Q)(\mu-a)}},$$

while, by construction, $\rho_0(\mu) = 1$, for $\mu \in [0, a]$, and $\rho_0(\mu) = 0$, for $\mu \in [b, R]$.

Requiring $W(z) \sim 1/z$, as $z \rightarrow \infty$, from (2.13) we obtain the following equations for the end-points a and b :

$$\frac{\sqrt{b+Q} - \sqrt{a+Q}}{\sqrt{b} + \sqrt{a}} = \sqrt{\alpha}, \quad \frac{\sqrt{(b+Q)(a+Q)} + \sqrt{ab} - Q}{2} = 1. \quad (2.15)$$

These equations can be easily solved, with the result (2.14)

It is clear from qualitative arguments that the current regime holds for relatively small values of Q , and relatively large values of R , so that the minimum of the potential is relatively close to the left hard wall at origin, and relatively far from the right hard wall at R .

Imposing $a = 0$, we get the value $Q = Q_c$ corresponding to the soft-edge/hard edge transition induced by the left hard wall at origin. We have

$$Q_c = \frac{1 - \alpha}{\alpha}. \quad (2.16)$$

It is clear that Regime IA corresponds to values of $Q < Q_c$.

Imposing $b = R$, we get the value $R = R_c$ corresponding to the soft-edge/hard edge transition induced by the right hard wall located at R . We have

$$R_c = \frac{(1 + \sqrt{\alpha(1+Q)})^2}{1 - \alpha}. \quad (2.17)$$

Clearly, this is the minimum value of R for which the scenario of Regime IA is applicable.

PROPOSITION 2.2. *The resolvent of the equilibrium measure for the continuum limit of the discrete Coulomb gas (1.6) in Regime IB is*

$$W(z) = -\log \sqrt{\alpha} + \log \frac{\sqrt{a(z-b)} + \sqrt{b(z-a)}}{\sqrt{(a+Q)(z-b)} + \sqrt{(b+Q)(z-a)}}. \quad (2.18)$$

where the parameters a and b , the two band end-points, are given again by (2.14)

In Regime IB, there is one band $[a, b]$ between two voids, $[0, a]$ and $[b, R]$. The resolvent $W(z)$ is simply given by the solution of (2.11) with $U(z) = V'(z)$ and $I_0 = [a, b]$. As a result, after evaluating the integral (2.12), we get expression (2.18).

The density in the band is given by

$$\rho_0(\mu) = -\frac{1}{\pi} \arctan \sqrt{\frac{a(b-\mu)}{b(\mu-a)}} + \frac{1}{\pi} \arctan \sqrt{\frac{(a+Q)(b-\mu)}{(b+Q)(\mu-a)}}, \quad \mu \in [a, b],$$

while, by construction, $\rho_0(\mu) = 0$, for $\mu \in [0, a] \cup [b, R]$.

Requiring $W(z) \sim 1/z$, as $z \rightarrow \infty$, from (2.18) we have the following equations for the end-points a and b :

$$\frac{\sqrt{b+Q} - \sqrt{a+Q}}{\sqrt{b} - \sqrt{a}} = \sqrt{\alpha}, \quad \frac{\sqrt{(a+Q)(b+Q)} - \sqrt{ab} - Q}{2} = 1.$$

Note that these equations differ from those in Regime IA, see (2.15), just by the sign of \sqrt{a} . Their formal solution, satisfying $0 < a < b$, is given again by (2.14), in which now $Q > Q_c$, see (2.16).

We also note that the critical value of R , namely, its minimal value, for which the Regime IB is applicable, is given by (2.17). Thus the minimum value of R for which the scenario of Regime IB is applicable is the same as for Regime IA.

In Regime IB, all charges accumulate in the vicinity of the minimum μ_0 , and one should check that the obtained solution for (2.11) with $U(z) = V'(z)$ and $I_0 = [a, b]$ satisfies condition (2.6). This is indeed the case, thus ensuring that no saturated region arises.

Summarizing, a critical value Q_c exists, that separates the two regimes, with Regime IA corresponding to $Q < Q_c$ and Regime IB to $Q > Q_c$. The value Q_c is determined by the condition $a = 0$, that is by the vanishing of the left gap, and the corresponding soft-edge/hard-edge transition. Similarly, a critical value R_c , determined by the condition $R = b$, exists, such that Regime IA and IB correspond to $R > R_c$. The value R_c is determined by the condition $b = R$, that is by the vanishing of the right gap, and the corresponding soft-edge/hard-edge transition.

2.4. Regime II. Regime IIA and Regime IIB are closely related. Actually, they can be viewed as two different scenarios for one same regime, Regime II, characterized by a saturated region on the right gap $[b, R]$. They both correspond to values $R < R_c$, with R_c given by (2.17).

PROPOSITION 2.3. *The resolvent of the equilibrium measure for the continuum limit of the discrete Coulomb gas (1.6) in Regime IIA is*

$$W(z) = -\log \sqrt{\alpha} + \log \frac{z}{z-R} - \log \frac{\sqrt{a(z-b)} + \sqrt{b(z-a)}}{\sqrt{(R-a)(z-b)} + \sqrt{(R-b)(z-a)}} - \log \frac{\sqrt{(a+Q)(z-b)} + \sqrt{(b+Q)(z-a)}}{\sqrt{(R-a)(z-b)} + \sqrt{(R-b)(z-a)}}. \quad (2.19)$$

where the parameters a and b , the two band end-points, are the solutions of the two equations:

$$\begin{aligned} \sqrt{\alpha} \frac{\sqrt{R-b} - \sqrt{R-a}}{\sqrt{R-b} + \sqrt{R-a}} \frac{\sqrt{b} + \sqrt{a}}{\sqrt{b+Q} - \sqrt{a+Q}} &= 1, \\ \frac{\sqrt{ab} + \sqrt{(a+Q)(b+Q)} - Q}{2} + \sqrt{(R-a)(R-b)} &= 1. \end{aligned} \quad (2.20)$$

In Regime IIA there is a band in $[a, b]$ between two saturated regions in $[0, a]$ and $[b, R]$. This implies that the resolvent in this case has the form

$$W(z) = \log \frac{z}{z-a} + \log \frac{z-b}{z-R} + H(z), \quad H(z) := \int_a^b \frac{\rho_0(\mu)}{z-\mu} d\mu.$$

The function $H(z)$ should solve equation (2.11), where the function $U(\mu)$ is given by the expression

$$U(\mu) = -\log \alpha + \log \frac{(\mu-a)^2(\mu-b)^2}{\mu(\mu+Q)(\mu-R)^2}.$$

Evaluating the integral (2.12), and coming back to function $W(z)$, we get (2.19).

The density is given by

$$\begin{aligned} \rho_0(\mu) = 1 + \frac{1}{\pi} \arctan \sqrt{\frac{a(b-\mu)}{b(\mu-a)}} + \frac{1}{\pi} \arctan \sqrt{\frac{(a+Q)(b-\mu)}{(b+Q)(\mu-a)}} \\ - \frac{2}{\pi} \arctan \sqrt{\frac{(R-a)(b-\mu)}{(R-b)(\mu-a)}}, \quad \mu \in [a, b], \end{aligned}$$

with $\rho_0(\mu) = 1$, for $\mu \in [0, a] \cup [b, R]$, by construction.

Imposing $W(z) \sim 1/z$, as $z \rightarrow \infty$, from (2.19) the system of equations (2.20) is obtained for the end-points a and b .

PROPOSITION 2.4. *The resolvent of the equilibrium measure for the continuum limit of the discrete Coulomb gas (1.6) in Regime IIB is*

$$W(z) = -\log \sqrt{\alpha} - 2 \log \frac{\sqrt{(b-a)(z-R)}}{\sqrt{(R-a)(z-b)} + \sqrt{(R-b)(z-a)}} + \log \frac{\sqrt{a(z-b)} + \sqrt{b(z-a)}}{\sqrt{(a+Q)(z-b)} + \sqrt{(b+Q)(z-a)}}. \quad (2.21)$$

where the parameters a and b , the two band end-points, are the solutions of the two equations:

$$\begin{aligned} \sqrt{\alpha} \frac{\sqrt{R-b} - \sqrt{R-a}}{\sqrt{R-b} + \sqrt{R-a}} \frac{\sqrt{b} - \sqrt{a}}{\sqrt{b+Q} - \sqrt{a+Q}} &= 1, \\ \frac{-\sqrt{ab} + \sqrt{(a+Q)(b+Q)} - Q}{2} + \sqrt{(R-a)(R-b)} &= 1. \end{aligned} \quad (2.22)$$

In Regime IIB there is a band in $[a, b]$, with a saturated region in $[b, R]$. Therefore the resolvent has the form

$$W(z) = \log \frac{z-b}{z-R} + H(z), \quad H(z) := \int_a^b \frac{\rho_0(\mu)}{z-\mu} d\mu.$$

The function $H(z)$ should solve equation (2.11), where the function $U(\mu)$ reads

$$U(\mu) = -\log \alpha + \log \frac{\mu(\mu-b)^2}{(\mu+Q)(\mu-R)^2}.$$

As a result, expression (2.21) is obtained for the resolvent.

The density is given by

$$\begin{aligned} \rho_0(\mu) &= \frac{1}{\pi} \arctan \sqrt{\frac{a(b-\mu)}{b(\mu-a)}} - \frac{1}{\pi} \arctan \sqrt{\frac{(a+Q)(b-\mu)}{(b+Q)(\mu-a)}} \\ &\quad - \frac{2}{\pi} \arctan \sqrt{\frac{(R-a)(b-\mu)}{(R-b)(\mu-a)}}, \quad \mu \in [a, b], \end{aligned}$$

with $\rho_0(\mu) = 0$, for $\mu \in [0, a]$, and $\rho_0(\mu) = 1$, for $\mu \in [b, R]$, by construction.

The condition that the resolvent (2.21) has the behaviour $W(z) \sim 1/z$, as $z \rightarrow \infty$, leads to the system of equations (2.22) for the end-points a and b .

Just as for the case of Regime I, Regime IIA and Regime IIB corresponds to values of Q in the ranges $Q < Q_c$ or $Q > Q_c$, respectively. The value Q_c is given by the condition that $a = 0$, that is by the vanishing of the left gap, and the corresponding soft-edge/hard-edge transition. In Regime II the critical value Q_c has not the simple form (2.16), but it is given instead by some more complicate expression involving the parameter R in a non-trivial way.

Finally, note that both in Regime IIA and Regime IIB, as the parameter R is decreased the band $[a, b]$ shrinks down, vanishing in the limit $R \rightarrow 1$. More precisely, in Regime IIA, one has $a \rightarrow b \neq 0$, and in Regime IIB $a, b \rightarrow 0$ as $R \rightarrow 1$. In both cases, the density tends to the saturated value 1 everywhere in the interval $[0, 1]$.

2.5. The Arctic ellipse. Regime I and Regime II discussed above correspond to values of R in the interval $R > R_c$, and $1 \leq R < R_c$, respectively, where $R_c = R_c(Q)$ is given by (2.17), and Q is allowed to take arbitrary values in the interval $[0, \infty)$. We discuss here how these conditions translate in terms of the coordinates $x, y \in [0, 1]$.

Clearly, the transition between the two regimes corresponds to $R = R_c(Q)$, that gives a relation between the two parameters R and Q , namely

$$\sqrt{(1-\alpha)R} = 1 + \sqrt{\alpha(1+Q)}.$$

Recalling (2.3), this relation translates into the following relation between the coordinates x and y :

$$\sqrt{y} = \sqrt{(1-\alpha)(1-x)} - \sqrt{\alpha x}. \quad (2.23)$$

This is readily recognized as the portion of the Arctic ellipses (1.9), between the two contact points $(x, y) = (1-\alpha, 0)$ and $(x, y) = (0, 1-\alpha)$. The additional condition $x \geq y$ (corresponding to the condition $R \geq 0$) selects the arc of ellipse \mathcal{A} , see (1.10), separating the two domains \mathcal{D}_I and \mathcal{D}_{II} , see Fig. 4.

It can also be easily seen that the values of x and y , which corresponds to $R > R_c$, i.e., to Regime I, are subject to the conditions $y < y_c$ and $y \leq x$, where $y_c = y_c(x)$ follows from (2.23), and is given by

$$y_c = \begin{cases} (\sqrt{(1-\alpha)(1-x)} - \sqrt{\alpha x})^2 & x \in [(1-\sqrt{\alpha})/2, 1-\alpha] \\ 0 & x \in [1-\alpha, 1]. \end{cases}$$

This conditions implies that in Regime I we have $(x, y) \in \mathcal{D}_I$. Similarly, one can verify that Regime II corresponds to values of the coordinates x and y such that $(x, y) \in \mathcal{D}_{II}$.

Summarizing, Regime I of the Coulomb gas corresponds in the six-vertex model in an L-shaped domain to situations where the cut-off corner is sufficiently small to lie totally outside the Arctic ellipse. Regime II corresponds in the six-vertex model to situations where the cut-off corner is sufficiently large to overlap with the Arctic ellipse. As explained in the next section, this transition manifests as a third-order phase transition in the free energy of the Coulomb gas, and consequently, of the six-vertex model in an L-shaped domain. Instead, the transition between Regime IA and Regime IB (or also between Regime IIA and Regime IIB), although being as well a soft-edge/hard-edge transition, does not induce any discontinuity in the Coulomb gas free energy.

3. The free energy

To evaluate function $\Phi(R, Q)$, we exploit the specific dependence of the discrete Coulomb gas (1.6) on the variable α . Recall the definition (2.8) of the first moment of the equilibrium measure. Clearly, we have the relation

$$E = \lim_{s \rightarrow \infty} \frac{1}{s^2} \frac{\partial \log I_{r,s,q}}{\partial \log \alpha}. \quad (3.1)$$

Given the resolvent $W(z)$, the quantity E can be easily extracted as the $1/z^2$ coefficient of its large z expansion, see (2.7). Using (2.1) in (3.1), we have the following ordinary first-order differential equation for the function $\Phi(R)$:

$$\frac{\partial}{\partial \log \alpha} \Phi(R, Q) = E.$$

This equation can be easily solved,

$$\Phi(R, Q) = \int E \frac{d\alpha}{\alpha} + C(R, Q). \quad (3.2)$$

where $C(R, Q)$ is some function independent of α .

To fix the integration constant $C(R, Q)$, we use the fact that the function $\varphi(x, y)$ admits direct evaluation when the parameter α takes the limiting values $\alpha = 0$ and $\alpha = 1$. These two cases play the role of initial conditions in determining the function $\varphi(x, y)$ in Regime I and Regime II, respectively.

3.1. The limiting cases. Let us first consider the case $\alpha = 0$. To start with, let us come back to the Coulomb gas at finite values of r, s, q . From (1.6) it follows that the expression $I_{r,s,q}$ is a polynomial in α . Its lowest order term is of degree $s(s-1)/2$; the corresponding coefficient can be found (modulo the number of permutations of m_j 's) by setting $m_j = j - 1$, that gives

$$I_{r,q,s} \sim \alpha^{s(s-1)/2}, \quad \alpha \rightarrow 0.$$

This, together with (1.5), implies

$$F_{r,s,q}|_{\alpha=0} = 1.$$

Hence, by (1.8), we have, in particular,

$$\varphi(x, y)|_{\alpha=0} = 0. \quad (3.3)$$

This condition will allow us to fix the integration constant $C(R, Q)$ in Regime I.

Let us now turn to the case $\alpha = 1$. Since the function $F_{r,s,q}$ has a zero at $\alpha = 1$ of degree $s(s+q)$, the function $\varphi(x, y) = \varphi(x, y; \alpha)$ has a logarithmic singularity as $\alpha \rightarrow 1$. Besides this, the function $\varphi(x, y)$ has a regular $\alpha \rightarrow 1$ part, which is relevant for our analysis.

To compute it, we focus on the Hankel determinant appearing in $F_{r,s,q}$, see representation (1.4). In the case $\alpha = 1$, the determinant can be evaluated in closed form.

PROPOSITION 3.1. *For arbitrary non-negative integers r, s, q , $s \leq r$, the following holds*

$$\det_{1 \leq j, k \leq s} \left[\sum_{m=0}^{r-1} m^{j+k-2} \binom{m+q}{m} \right] = \prod_{j=0}^{s-1} \frac{(j!(j+q)!)^2 (j+q+r)!}{q!(r-j-1)!(2j+q)!(2j+q+1)!}.$$

Indeed, the determinant in the left hand side can be recognized as the Gram determinant of Hahn polynomials $Q_j(m; q, 0, r-1)$. The statement follows from the standard technique of orthogonal polynomials. For an account of the properties of Hahn polynomials, see, e.g., the monograph [31], Sect. 9.5.

Therefore, as $\alpha \rightarrow 1$,

$$F_{r,s,q} \sim \prod_{j=0}^{s-1} \frac{j!(j+q)!(j+q+r)!}{q!(r-j-1)!(2j+q)!(2j+q+1)!} \cdot (1-\alpha)^{s(s+q)}. \quad (3.4)$$

Let $s+q = [xN]$, $s = [yN]$, with $x, y \in [0, 1]$ fixed, $N = r+s+q$. We set

$$\psi(x, y) = - \lim_{N \rightarrow \infty} \frac{1}{N^2} \log \left(\prod_{j=0}^{s-1} \frac{j!(j+q)!(j+q+r)!}{q!(r-j-1)!(2j+q)!(2j+q+1)!} \right).$$

From (3.4), recalling (1.8), it follows that

$$\lim_{\alpha \rightarrow 1} [\varphi(x, y) + xy \log(1 - \alpha)] = \psi(x, y). \quad (3.5)$$

Using the Stirling formula, we have

$$\begin{aligned} \psi(x, y) = & \frac{(x+y)^2}{2} \log(x+y) - \frac{(1-x-y)^2}{2} \log(1-x-y) \\ & - \frac{x^2}{2} \log x + \frac{(1-x)^2}{2} \log(1-x) - \frac{y^2}{2} \log y + \frac{(1-y)^2}{2} \log(1-y). \end{aligned} \quad (3.6)$$

Recalling (2.2), relation (3.5) implies

$$\Phi(R, Q)|_{\alpha=1} = -(R+Q+1)^2 \psi(x, y).$$

Recall that x and y are related to R and Q by

$$x = \frac{1+Q}{R+Q+1}, \quad y = \frac{1}{R+Q+1},$$

see (2.3). Finally, due to (3.6), we obtain the condition

$$\begin{aligned} \Phi(R, Q)|_{\alpha=1} = & -\frac{R^2}{2} \log R + \frac{(R-1)^2}{2} \log(R-1) - \frac{(R+Q)^2}{2} \log(R+Q) \\ & + \frac{(1+Q)^2}{2} \log(1+Q) - \frac{(2+Q)^2}{2} \log(2+Q) \\ & + \frac{(R+Q+1)^2}{2} \log(R+Q+1). \end{aligned} \quad (3.7)$$

This condition make it possible to fix the integration constant $C(R, Q)$ in Regime II.

After these preliminary remarks, we turn to the evaluation of the free energy of the discrete Coulomb gas in both regimes.

3.2. Regime I. In order to extract the first moment of the equilibrium measure in Regime IA and Regime IB, we evaluate the term $O(1/z^2)$ in the $z \rightarrow \infty$ expansion of the resolvent, see (2.13) and (2.18), respectively. In both regimes we obtain

$$E = \frac{2+Q}{8}(a+b) + \frac{Q^2}{4} - \frac{Q}{4} \sqrt{(a+Q)(b+Q)}.$$

Substituting here the expressions for the end-points a and b , see (2.14), we have

$$E = (Q+1) \frac{\alpha}{1-\alpha} + \frac{1}{2}.$$

Integrating with respect to α , and recalling (3.2), we get the free energy of the discrete Coulomb gas for Regime I,

$$\Phi(R, Q) = -(Q+1) \log(1-\alpha) + \log \sqrt{\alpha} + C(R, Q),$$

where the integration constant $C(R, Q)$ is still to be determined. The last relation implies

$$\varphi(x, y) = C(R, Q),$$

see (2.2). Finally, the initial condition (3.3) determines the value of the integration constant,

$$C(R, Q) = 0.$$

As a result, we obtain

$$\varphi(x, y) = 0, \quad (x, y) \in \mathcal{D}_I,$$

as stated in Theorem 1.4.

3.3. Regime II: The end-point equations. In the previous section we obtained that the end-points a and b of the band in Regime IIA, see (2.20), and Regime IIB, see (2.22), obey the equations

$$\begin{aligned} \sqrt{a} \frac{\sqrt{R-b} - \sqrt{R-a}}{\sqrt{R-b} + \sqrt{R-a}} \frac{\sqrt{b} + \nu\sqrt{a}}{\sqrt{b+Q} - \sqrt{a+Q}} &= 1, \\ \frac{\nu\sqrt{ab} + \sqrt{(a+Q)(b+Q)} - Q}{2} + \sqrt{(R-a)(R-b)} &= 1, \end{aligned} \quad (3.8)$$

where

$$\nu = \begin{cases} +1 & \text{for Regime IIA} \\ -1 & \text{for Regime IIB.} \end{cases}$$

These are essentially equivalent to quartic equations; a direct treatment is given in appendix A. The most important concern about the end-points in the approach we follow here, is that they need to be represented in such a way that the integral in (3.2) might be evaluated explicitly.

For this reason we consider here a solution of equations (3.8) in which the second equation is viewed as defining an algebraic curve in the variables a and b , while the first equation is regarded as fixing a point on this curve. Indeed, by squaring the second equation twice (since it involves the linear combination of three square roots with different arguments), a and b are easily seen to lie on an algebraic curve of degree four. It turns out that in suitable variables, this is simply a cubic curve, which is moreover of rational (rather than elliptic) type. In this construction the first equation in (3.8) becomes a quartic equation for the parameter along this curve. Furthermore, the solution of this quartic equation can be viewed as a change of variables, so that the integration (3.2) turns out to be feasible.

Instead of dealing with the end-points a and b as unknowns, we introduce the quantities

$$\begin{aligned} A_{\pm} &= \frac{(\sqrt{b} \pm \nu\sqrt{a})^2}{4}, \\ B_{\pm} &= \frac{(\sqrt{b+Q} \pm \sqrt{a+Q})^2}{4}, \\ C_{\pm} &= \frac{(\sqrt{R-a} \pm \sqrt{R-b})^2}{4}. \end{aligned} \quad (3.9)$$

Obviously, the end-points can be expressed in terms of these new quantities, for example

$$a = A_+ + A_- - 2\sqrt{A_+A_-}, \quad b = A_+ + A_- + 2\sqrt{A_+A_-}.$$

Similar expressions are valid also in terms of B 's and C 's. Namely, the quantities (3.9) satisfy the multiplicative self-consistency relations

$$A_+A_- = B_+B_- = C_+C_-, \quad (3.10)$$

together with the additive ones

$$A_+ + A_- = B_+ + B_- - Q = R - C_+ - C_-. \quad (3.11)$$

The first equation in (3.8) now reads

$$\alpha \frac{C_- A_+}{C_+ B_-} = 1, \quad (3.12)$$

and the second equation can be written in either of two forms

$$A_{\pm} + B_{\pm} + 2C_{\pm} = N_{\pm}, \quad (3.13)$$

where

$$N_+ := R + Q + 1, \quad N_- := R - 1.$$

Obviously, (3.10), (3.11), (3.12), and (3.13) constitute a system of six equations for the six unknowns A_{\pm} , B_{\pm} , and C_{\pm} .

Before turning to solution of this system, it is useful to mention that starting from one form of equation (3.13), say, using the ‘+’ sign, the other form, with the ‘−’ sign, can be directly found from (3.11). Indeed, rewriting (3.11) as

$$\begin{aligned} A_+ + A_- + C_+ + C_- &= R, \\ B_+ + B_- + C_+ + C_- &= R + Q, \end{aligned} \quad (3.14)$$

and summing these equations, it is clear that the two forms in equation (3.13) are equivalent.

Dividing the two equations in (3.13) by $\sqrt{A_{\pm} B_{\pm}}$, respectively, and using the relation $A_+/B_+ = B_-/A_-$, which follows from the first equality in (3.10), a comparison of the resulting relations yields

$$\frac{N_+ - 2C_+}{\sqrt{A_+ B_+}} = \frac{N_- - 2C_-}{\sqrt{A_- B_-}}.$$

Therefore, using (3.10), we can readily express A ’s in terms of B ’s and C ’s, as follows:

$$A_{\pm} = B_{\mp} \frac{z_{\pm}}{z_{\mp}}, \quad (3.15)$$

where we have used the notation

$$z_{\pm} := N_{\pm} - 2C_{\pm}.$$

We may now use these expressions in (3.14) to eliminate the B ’s.

Using (3.15), and writing C ’s in terms of z ’s, from (3.14) we have

$$\begin{aligned} B_+ \frac{z_-}{z_+} + B_- \frac{z_+}{z_-} &= \frac{z_+ + z_-}{2} - \frac{Q}{2}, \\ B_+ + B_- &= \frac{z_+ + z_-}{2} + \frac{Q}{2}. \end{aligned}$$

Solving for B ’s, we obtain

$$B_{\pm} = \left(1 \pm \frac{Q}{z_+ - z_-} \right) \frac{z_{\pm}}{2}. \quad (3.16)$$

Recalling that $B_+ B_- = C_+ C_-$, see (3.10), we have

$$\left(1 - \frac{Q^2}{(z_+ - z_-)^2} \right) z_+ z_- = (N_+ - z_+)(N_- - z_-).$$

Simplifying the $z_+ z_-$ term standing in both sides, we obtain the cubic equation:

$$Q^2 z_+ z_- = (N_- z_+ + N_+ z_- - N_+ N_-)(z_+ - z_-)^2. \quad (3.17)$$

Note that it contains only the third and second order terms in z 's, thus describing a rational curve. This means that z 's can be expressed as a rational function of some parameter along the curve.

To obtain this parametrization, we set $z_+ = tz_-$ in (3.17) and consider z_- as a function of the parameter t . Then we get

$$z_- = \frac{1}{N_-t + N_+} \left[\frac{Q^2t}{(t-1)^2} + N_+N_- \right], \quad z_+ = tz_-.$$

Note that we need just that portion of this curve for which the C 's are positive. It can be easily seen that this corresponds to values of the parameter t in the interval $[N_+/N_-, \infty)$. It is useful to introduce a new parameter $\eta \in [0, 1]$, which parametrizes this portion of the curve as follows:

$$t = \frac{N_+}{N_-} \frac{1+\eta}{1-\eta}.$$

In terms of the new parameter, z 's read

$$z_{\pm} = N_{\pm} \frac{1 \pm \eta}{2} \left(\frac{Q^2(1-\eta)(1+\eta)}{(N_+(1+\eta) - N_-(1-\eta))^2} + 1 \right). \quad (3.18)$$

As a result, recalling that $C_{\pm} = (E_{\pm} - z_{\pm})/2$, using expressions for B 's (3.16), and those for A 's (3.15), we obtain that the parametrization (3.18) implies nice factorized expressions:

$$\begin{aligned} C_+ &= N_+ \frac{(1-\eta)(1+R\eta)(1+Q+(R+Q)\eta)}{(2+Q+(2R+Q)\eta)^2}, \\ C_- &= N_- \frac{(1+\eta)(1+Q+R\eta)(1+(R+Q)\eta)}{(2+Q+(2R+Q)\eta)^2}, \\ B_+ &= N_+ \frac{(1+\eta)(1+Q+R\eta)(1+Q+(R+Q)\eta)}{(2+Q+(2R+Q)\eta)^2}, \\ B_- &= N_- \frac{(1-\eta)(1+R\eta)(1+(R+Q)\eta)}{(2+Q+(2R+Q)\eta)^2}, \\ A_+ &= N_+ \frac{(1+\eta)(1+R\eta)(1+(R+Q)\eta)}{(2+Q+(2R+Q)\eta)^2}, \\ A_- &= N_- \frac{(1-\eta)(1+Q+R\eta)(1+Q+(R+Q)\eta)}{(2+Q+(2R+Q)\eta)^2}. \end{aligned} \quad (3.19)$$

These expressions solve equations (3.10), (3.11), and (3.13), and moreover they correspond to the solution in which all A 's, B 's, and C 's are positive, provided that $\eta \in [0, 1]$.

Note that till now equation (3.12) has never been used yet. To obtain the solution of the initial end-points problem, we exploit this last equation. Using (3.19), we have

$$\alpha \frac{(1+\eta)^2(1+Q+R\eta)(1+(R+Q)\eta)}{(1-\eta)^2(1+R\eta)(1+Q+(R+Q)\eta)} = 1. \quad (3.20)$$

This is a quartic equation, and we need only that root taking values in the interval $[0, 1]$. It can be verified that such a root $\eta = \eta(R, Q; \alpha)$ always exists, provided that the parameters R , Q , and α correspond to Regime II. An explicit expression for this root is not crucial for what follows; it is discussed in Appendix A.

3.4. Regime II: The free energy. In order to extract the first moment of the equilibrium measure in Regime IIA and Regime IIB, we evaluate the term $O(1/z^2)$ in the $z \rightarrow \infty$ expansion of the resolvent, see (2.19) and (2.21), respectively. In both regimes we obtain

$$E = \frac{2+Q}{8}(a+b) + \frac{Q^2}{4} - \frac{Q}{4}\sqrt{(a+Q)(b+Q)} + \frac{R}{2}\sqrt{(R-a)(R-b)}.$$

Expressing E in terms of quantities (3.9), we have

$$E = \frac{Q^2}{4} + \frac{2+Q}{4}(A_+ + A_-) - \frac{Q}{4}(B_+ - B_-) + \frac{R}{2}(C_+ - C_-).$$

To obtain the function $\Phi(R, Q)$, we write the integral in (3.2) as

$$\Phi(R, Q) = \int E \frac{\partial \log \alpha}{\partial \eta} d\eta + C(R, Q),$$

where the functions $E = E(\eta)$ and $\alpha = \alpha(\eta)$ follow from (3.19) and (3.20), respectively.

Evaluating the integral, we find that

$$\Phi(R, Q) = \Omega(R, Q) + C(R, Q), \quad (3.21)$$

where the function $\Omega(R, Q) = \Omega(R, Q; \eta)$ is given as a linear combination of seven logarithms:

$$\begin{aligned} \Omega(R, Q) = & \left(\frac{1}{2} + R - \frac{R^2}{2} \right) \log(1 - \eta) + \left(\frac{1}{2} - R + Q - \frac{(R+Q)^2}{2} \right) \log(1 + \eta) \\ & + \left(\frac{1}{2} + R \right) \log(1 + R\eta) + \left(\frac{1}{2} - R + Q - RQ \right) \log(1 + Q + R\eta) \\ & + \left(\frac{1}{2} - R \right) \log(1 + (R+Q)\eta) \\ & + \left(\frac{1}{2} + R + Q + RQ + Q^2 \right) \log(1 + Q + (R+Q)\eta) \\ & - \frac{(2+Q)^2}{2} \log(2 + Q + (2R+Q)\eta). \end{aligned} \quad (3.22)$$

To find the integration constant $C(R, Q)$, we recall that in our preliminary discussion we obtained the condition (3.7). To employ this condition, we note that from (3.20) it follows that $\eta \rightarrow 0$ as $\alpha \rightarrow 1$, hence the constant $C(R, Q)$ can be inferred by equating (3.21) at $\eta = 0$ with the expression (3.7), that gives

$$\begin{aligned} C(R, Q) = & -\frac{R^2}{2} \log R + \frac{(R-1)^2}{2} \log(R-1) - \frac{(R+Q)^2}{2} \log(R+Q) \\ & - \frac{(1+Q)^2}{2} \log(1+Q) + \frac{(R+Q+1)^2}{2} \log(R+Q+1). \end{aligned} \quad (3.23)$$

Expressions (3.21), (3.22), and (3.23) provide a solution of the free energy problem of the discrete Coulomb gas in Regime II.

From (2.2) and (2.3), it follows that in Regime II the function $\varphi(x, y)$ is given by the following expression:

$$\varphi(x, y) = \chi(x, y; \eta) - \chi_0(x, y), \quad (x, y) \in \mathcal{D}_{\text{II}}, \quad (3.24)$$

where

$$\begin{aligned} \chi(x, y; \eta) = & -xy \log(1 - \alpha) + \frac{y^2}{2} \log \alpha \\ & + \left(\frac{x^2}{2} + xy - \frac{y^2}{2} - x - y + \frac{1}{2} \right) \log(1 - \eta) + \left(y^2 - 2xy + \frac{1}{2} \right) \log(1 + \eta) \\ & - \left(\frac{y}{2} - x + 1 \right) y \log[y + (1 - x)\eta] - \left(x^2 + xy - \frac{y^2}{2} - x \right) \log[x + (1 - x)\eta] \\ & - \left(\frac{y}{2} + x - 1 \right) y \log[y + (1 - y)\eta] - \left(\frac{y^2}{2} - xy + x \right) \log[x + (1 - y)\eta], \\ & + \frac{(x + y)^2}{2} \log[x + y + (2 - x - y)\eta], \quad (3.25) \end{aligned}$$

and the function $\chi_0(x, y)$, independent of η , reads

$$\begin{aligned} \chi_0(x, y) = & -\frac{x^2}{2} \log x - \frac{(1 - x)^2}{2} \log(1 - x) - \frac{y^2}{2} \log y - \frac{(1 - y)^2}{2} \log(1 - y) \\ & + \frac{(1 - x - y)^2}{2} \log(1 - x - y). \end{aligned}$$

Taking into account that the equation for the parameter η in terms of variables x and y reads

$$\alpha \frac{(1 + \eta)^2 [x + (1 - x)\eta] [y + (1 - y)\eta]}{(1 - \eta)^2 [y + (1 - x)\eta] [x + (1 - y)\eta]} = 1 \quad (3.26)$$

and also using that

$$1 - \alpha = \frac{\eta [x + y + (2 - x - y)\eta]^2}{(1 + \eta)^2 [x + (1 - x)\eta] [y + (1 - y)\eta]},$$

we may express α in terms of η in (3.25), thus obtaining for $\chi(x, y; \eta)$ the following expression:

$$\begin{aligned} \chi(x, y; \eta) = & -xy \log \eta + \frac{(1 - x - y)^2}{2} \log(1 - \eta) + \frac{1}{2} \log(1 + \eta) \\ & - (1 - x)y \log[y + (1 - x)\eta] - x(1 - y) \log[x + (1 - y)\eta] \\ & + x(1 - x) \log[x + (1 - x)\eta] + y(1 - y) \log[y + (1 - y)\eta] \\ & + \frac{(x - y)^2}{2} \log[x + y + (2 - x - y)\eta]. \quad (3.27) \end{aligned}$$

We conclude by stating a remarkable property of representation (3.24).

PROPOSITION 3.2. *Let:*

$$h = h(x, y) := \sqrt{\frac{xy}{(1 - x)(1 - y)}}.$$

Then the following holds:

$$\chi_0(x, y) = \chi(x, y; h). \quad (3.28)$$

Verification is fairly straightforward. Representation (3.24) together with (3.27) and (3.28) immediately implies formula (1.11), which is thus proven.

The proofs of Propositions 1.7 and 1.8 are based on specific properties of the function $\chi(x, y; \eta)$, namely that its first and second derivative with respect to η vanish at $\eta = h$. Further details and explicit calculations are given in Appendix B.

Appendix A. End-point equations in Regime II: A direct treatment

We discuss here a rather direct approach to treat the end-point equations (2.20). Using notations (3.9), we note that these equations can be written using only A_+ , B_+ , and C_+ , or only A_- , B_- , and C_- ,

$$\frac{C_{\pm}}{\sqrt{A_{\pm}B_{\pm}}} = (\sqrt{\alpha})^{\pm 1},$$

$$A_{\pm} + B_{\pm} + 2C_{\pm} = N_{\pm}.$$

Introduce parameter γ by

$$\gamma := \frac{B_-}{A_-} = \frac{A_+}{B_+}. \quad (\text{A.1})$$

From the first and second way of writing of the end-point equations, we have, respectively,

$$A_+ = \frac{\gamma N_+}{1 + \gamma + 2\sqrt{\alpha\gamma}}, \quad A_- = \frac{\sqrt{\alpha}N_-}{(1 + \gamma)\sqrt{\alpha} + 2\sqrt{\gamma}}.$$

From the definition of the quantities A_{\pm} we have

$$a = A_+ + A_- - 2\sqrt{A_+A_-}, \quad b = A_+ + A_- + 2\sqrt{A_+A_-},$$

and

$$\sqrt{a+Q} = \sqrt{\frac{A_+}{\gamma}} - \sqrt{\gamma A_-}, \quad \sqrt{b+Q} = \sqrt{\frac{A_+}{\gamma}} + \sqrt{\gamma A_-}.$$

The expressions for the end-points a and b which follow from these two ways of writing are compatible with each other, provided that

$$Q + A_+ + A_- = \frac{A_+}{\gamma} + \gamma A_-.$$

Explicitly, this is an equation for the parameter γ ,

$$Q + (1 - \gamma) \frac{\sqrt{\alpha}N_-}{(1 + \gamma)\sqrt{\alpha} + 2\sqrt{\gamma}} = (1 - \gamma) \frac{N_+}{1 + \gamma + 2\sqrt{\alpha\gamma}}.$$

Dividing by the expression standing in the right-hand side,

$$\frac{Q}{N_+} \frac{1 + \gamma + 2\sqrt{\alpha\gamma}}{1 - \gamma} + \frac{\sqrt{\alpha}N_-}{N_+} \frac{1 + \gamma + 2\sqrt{\alpha\gamma}}{(1 + \gamma)\sqrt{\alpha} + 2\sqrt{\gamma}} = 1,$$

and introducing a new parameter ω such that

$$\cos \omega = \frac{(1 + \gamma)\sqrt{\alpha} + 2\sqrt{\gamma}}{1 + \gamma + 2\sqrt{\alpha\gamma}}, \quad \sin \omega = \frac{\sqrt{1 - \alpha}(1 - \gamma)}{1 + \gamma + 2\sqrt{\alpha\gamma}}, \quad (\text{A.2})$$

we may rewrite the equation in the form

$$\frac{A}{\cos \omega} + \frac{B}{\sin \omega} = 1, \quad (\text{A.3})$$

where

$$A = \frac{\sqrt{\alpha}N_-}{N_+}, \quad B = \frac{\sqrt{1-\alpha}Q}{N_+}. \quad (\text{A.4})$$

It is interesting to note that the three parameters (α , R , and Q) in fact enters the equation only through their two combinations, A and B .

It is possible to find the roots of (A.3), by noting that it can be written as a quartic equation

$$x^4 - 2Bx^3 - (1 - A^2 - B^2)x^2 + 2Bx - B^2 = 0, \quad x = \sin \omega.$$

The roots of this equation are

$$x = \frac{B}{2} + \frac{1}{2} \left\{ \sqrt{B^2 + 2S \left(1 - \cos \frac{\theta}{3} \right)} \pm \sqrt{B^2 + 2S \left(1 - \cos \frac{\theta + 2\pi}{3} \right)} \right. \\ \left. \pm \sqrt{B^2 + 2S \left(1 - \cos \frac{\theta - 2\pi}{3} \right)} \right\}, \quad (\text{A.5})$$

where

$$\theta \in [0, \pi), \quad \theta = \arccos \left(1 - \frac{2A^2B^2}{S^3} \right), \quad S = \frac{1 - A^2 - B^2}{3} > 0,$$

and the signs are to be chosen independently. The root which lies in the interval $[0, 1]$ corresponds to the first and second signs in (A.5) being ‘+’ and ‘-’, respectively, as it can be checked by expanding the expression in (A.5) in power series, say, in B , and noting that the required root has the behaviour $x \sim (1 - A)^{-1}B$, as $B \rightarrow 0$.

To find a relation between η and parameter φ , we first express parameter γ , (A.1), in terms of η using (3.19),

$$\gamma = \frac{(1 + R\eta)(1 + (R + Q)\eta)}{(1 + Q + R\eta)(1 + Q + (R + Q)\eta)}.$$

Using (3.20) we also have

$$\sqrt{\alpha}\gamma = \frac{(1 - \eta)(1 + R\eta)}{(1 + \eta)(1 + Q + R\eta)}, \quad \sqrt{\frac{\gamma}{\alpha}} = \frac{(1 + \eta)(1 - R\eta)}{(1 + (R + Q)\eta)(1 + Q + (R + Q)\eta)},$$

and substituting in the expression for, say, $\cos \omega$, see (A.2), we obtain

$$\cos \omega = \sqrt{\alpha} \frac{1 + \eta}{1 - \eta}.$$

Consequently,

$$\eta = \frac{\cos \omega - \sqrt{\alpha}}{\cos \omega + \sqrt{\alpha}}. \quad (\text{A.6})$$

An expression involving $\sin \omega$ can be given by taking into account (A.3). Thus (A.5), with the indicated choice of the signs, provides an explicit expression for η .

Appendix B. Phase transition at the Arctic ellipse

Here we outline some calculations related to the third-order phase transition at the curve \mathcal{A} separating the regions \mathcal{D}_I and \mathcal{D}_{II} , see Fig. 4.

We first mention a particular parametrization of the points of \mathcal{A} which appears useful in these calculations. Recalling that \mathcal{A} is a portion of the ellipse

$$\frac{(1-x-y)^2}{\alpha} + \frac{(x-y)^2}{1-\alpha} = 0,$$

one can introduce an angle parameter ϕ , such that

$$1-x-y = \sqrt{\alpha} \sin \phi, \quad x-y = \sqrt{1-\alpha} \cos \phi.$$

If we further set $\alpha = \sin^2 \lambda$, $\lambda \in (0, \pi)$, then we have

$$x = \cos^2 \left(\frac{\phi + \lambda}{2} \right), \quad y = \sin^2 \left(\frac{\phi - \lambda}{2} \right), \quad (x, y) \in \mathcal{A}, \quad \phi \in [\lambda, \pi/2]. \quad (\text{B.1})$$

To illustrate the convenience of this parametrization, let us consider our quartic equation in the case of $(x, y) \in \mathcal{A}$. We take it in its ω -parametrized form (A.3). Recalling relations (2.3), for the coefficients A and B , see (A.4), we have $A = \sqrt{\alpha}(1-x-y)$ and $B = \sqrt{1-\alpha}(x-y)$, and so (A.3) in this case reads

$$\frac{\sin^2 \lambda \sin \phi}{\cos \omega} + \frac{\cos^2 \lambda \cos \phi}{\sin \omega} = 1.$$

The root we need is simply $\omega = \pi/2 - \phi$ (note that $\phi \in [\lambda, \pi/2]$ and $\lambda \in (0, \pi/2)$, and hence $\omega \in [0, \pi/2)$, as required). Substituting the obtained expression for the root in (A.6), we get

$$\eta|_{(x,y) \in \mathcal{A}} = \frac{\sin \phi - \sin \lambda}{\sin \phi + \sin \lambda} = \cot \left(\frac{\phi + \lambda}{2} \right) \tan \left(\frac{\phi - \lambda}{2} \right) = \sqrt{\frac{xy}{(1-x)(1-y)}}.$$

In the last expression one can recognize the quantity h , see (1.12), and hence its interpretation as the value of η on the Arctic ellipse.

Let us now turn to study the properties of the function $\varphi(x, y) = \varphi(x, y; \alpha)$ in the vicinity of \mathcal{A} . In particular, we want to verify that

$$\varphi(x, y) \sim C(\alpha - \alpha_c)^3, \quad \alpha \rightarrow \alpha_c^+.$$

Since $\varphi(x, y) = \chi(x, y; \eta) - \chi(x, y; h)$ where the function $\chi(x, y; \eta)$ is given by (3.27), and since the dependence on α enters only through η , the problem reduces to studying properties of $\chi(x, y; \eta)$ as a function η . By a direct calculation it can be verified that

$$\partial_\eta^k \chi(x, y; \eta)|_{\eta=h} = 0, \quad k = 1, 2. \quad (\text{B.2})$$

At the same time the third-order derivative does not vanish, that gives a non-zero value of $C = C(x, y)$, which can be computed by the formula

$$C = \partial_\alpha^3 \varphi|_{\alpha=\alpha_c} = \frac{1}{\alpha_c^3} \left[\frac{1}{(\partial_\eta \log \alpha)^3} \partial_\eta^3 \chi(x, y; \eta) \right] \Big|_{\eta=h},$$

where the function $\alpha = \alpha(\eta)$ is determined by (3.26). By a straightforward calculation we obtain

$$\begin{aligned} \partial_\eta \log \alpha|_{\eta=h} &= -\frac{2}{1-h} - \frac{2}{1+h} + \frac{1-x}{y+(1-x)h} + \frac{1-y}{x+(1-y)h} \\ &\quad - \frac{1-x}{x+(1-x)h} - \frac{1-y}{y+(1-y)h} = -4 \frac{(1-x)(1-y)}{1-x-y}. \end{aligned}$$

However, in the case of $\partial_\eta^3 \chi(x, y; \eta)$ the corresponding calculation appears to be much more involved; the parametrization (B.1) turns out to be very convenient (in particular, when resorting to computer symbolic calculations), providing the result in a nice factorized form:

$$\partial_\eta^3 \chi(x, y; \eta)|_{\eta=h} = -\frac{(\sin \phi - \sin \lambda)(\sin \phi + \sin \lambda)^7}{8 \sin \lambda \cos^4 \lambda \sin^3 \phi}.$$

Taking into account that $\partial_\eta \log \alpha|_{\eta=h} = -(\sin \phi + \sin \lambda)^2 / \sin \lambda \sin \phi$ and $\alpha_c = \sin^2 \lambda$, we finally obtain

$$C = \frac{2 \sin(\phi - \lambda) \sin(\phi + \lambda)}{\sin^4 2\lambda} = \frac{\sqrt{x(1-x)y(1-y)}}{2\alpha_c^2(1-\alpha_c)^2},$$

and we recall that α_c as a function of x and y is given by (1.15).

Similarly, the third-order phase transition can be regarded as a property of the function $\varphi(x, y) = \varphi(x, y; \alpha)$ with respect to the variables x and y , at fixed α . Namely, the function $\varphi(x, y) = \varphi(x, y; \alpha)$ has a discontinuity in its third normal derivative at the curve \mathcal{A} , with vanishing first and second normal derivative. Let t be the variable along the normal of the curve \mathcal{A} at the point (x, y) , then

$$\partial_t^k \varphi(x, y)|_{(x,y) \in \mathcal{A}} = 0, \quad k = 1, 2. \quad (\text{B.3})$$

Since $\varphi(x, y) = \chi(x, y; \eta) - \chi(x, y; h)$, in computing the derivatives with respect to t one has to take into account only those terms which involve derivatives of the functions $\chi(x, y; \eta)$ and $\chi(x, y; h)$ with respect to their parameters, η and h , respectively. Hence the property (B.3) is just a consequence of the analogous property for the function $\chi(x, y; \eta)$, see (B.2). For the third derivative we have

$$\partial_t^3 \varphi(x, y)|_{(x,y) \in \mathcal{A}} = [(\partial_t \eta|_{\eta=h})^3 - (\partial_t h)^3] \partial_\eta^3 \chi(x, y; \eta)|_{\eta=h}.$$

Since $\partial_t \eta|_{\eta=h}$ in general is not equal to $\partial_t h$, the third derivative is non-vanishing; its explicit expression can be obtained along the same lines as for the quantity C above.

References

- [1] D. Gensburg, I. Carlsen, and H.-Chr. Zapp, *Some exact results for the dimer problem on plane lattices with non-standard boundaries*, Phil. Mag. A **41** (1980), 777–781.
- [2] N. Elkies, G. Kuperberg, M. Larsen, and J. Propp, *Alternating-sign matrices and domino tilings*, J. Algebraic Combin. **1** (1992), 111–132; 219–234, arXiv:math/9201305.
- [3] W. Jockusch, J. Propp, and P. Shor, *Random domino tilings and the arctic circle theorem*, arXiv:math/9801068.
- [4] H. Cohn, M. Larsen, and J. Propp, *The shape of a typical boxed plane partition*, New York J. Math. **4** (1998), 137–165, arXiv:math/9801059.
- [5] R. Kenyon and A. Okounkov, *Limit shapes and the complex Burgers equation*, Acta Math. **199** (2007), 263–302, arXiv:math-ph/0507007.
- [6] R. Kenyon, A. Okounkov, and S. Sheffield, *Dimers and amoebae*, Ann. of Math. **163** (2006), 1019–1056, arXiv:math-ph/0311005.

- [7] R. Kenyon and A. Okounkov, *Planar dimers and Harnack curves*, Duke Math. J. **131** (2006), 499–524, arXiv:math-ph/0311005.
- [8] R. Kenyon, *Lectures on dimers*, IAS/Park City Math. Ser **16** (2009), 191–230, arXiv:0910.3129.
- [9] V. E. Korepin and P. Zinn-Justin, *Thermodynamic limit of the six-vertex model with domain wall boundary conditions*, J. Phys. A **33** (2000), 7053–7066, arXiv:cond-mat/0004250.
- [10] N. Reshetikhin and K. Palamarchuk, *The 6-vertex model with fixed boundary conditions*, PoS Solvay (2006), 012, arXiv:cond-mat/0502314.
- [11] F. Colomo and A. G. Pronko, *The arctic curve of the domain-wall six-vertex model*, J. Stat. Phys. **138** (2010), 662–700, arXiv:0907.1264.
- [12] F. Colomo and A. G. Pronko, *Third-order phase transition in random tilings*, Phys. Rev. E **88** (2013), 042125, arXiv:1306.6207.
- [13] K. Johansson, *Shape fluctuations and random matrices*, Comm. Math. Phys. **209** (2000), 437–476, arXiv:math/9903134.
- [14] K. Johansson, *Non-intersecting paths, random tilings, and random matrices*, Probab. Theory Related Fields **123** (2002), 225–280, arXiv:math/0011250.
- [15] P. Zinn-Justin, *Six-vertex model with domain wall boundary conditions and one-matrix model*, Phys. Rev. E **62** (2000), 3411–3418, arXiv:math-ph/0005008.
- [16] P. Bleher and K. Liechty, *Random matrices and the six vertex model*, CRM Monograph Series, vol. 32, American Mathematical Society, Providence, RI, 2013.
- [17] J. Baik, T. Kriecherbauer, K. T.-R. McLaughlin, and P. D. Miller, *Discrete orthogonal polynomials: Asymptotics and applications*, Ann. of Math. Stud., vol. 164, Princeton University Press, Princeton, NJ, 2007.
- [18] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics*, Academic Press, San Diego, CA, 1982.
- [19] V. E. Korepin, *Calculations of norms of Bethe wave functions*, Comm. Math. Phys. **86** (1982), 391–418.
- [20] A. G. Izergin, *Partition function of the six-vertex model in the finite volume*, Sov. Phys. Dokl. **32** (1987), 878–879.
- [21] A. G. Izergin, D. A. Coker, and V. E. Korepin, *Determinant formula for the six-vertex model*, J. Phys. A **25** (1992), 4315–4334.
- [22] N. M. Bogoliubov, A. G. Pronko, and M. B. Zvonarev, *Boundary correlation functions of the six-vertex model*, J. Phys. A **35** (2002), 5525–5541, arXiv:math-ph/0203025.
- [23] F. Colomo and A. G. Pronko, *Emptiness formation probability in the domain-wall six-vertex model*, Nucl. Phys. B **798** (2008), 340–362, arXiv:0712.1524.
- [24] A. G. Pronko, *On the emptiness formation probability in the free-fermion six-vertex model with domain wall boundary conditions*, J. Math. Sci. (N. Y.) **192** (2013), 101–116.
- [25] P. J. Forrester and N. S. Witte, *Application of the τ -function theory of Painlevé equations to random matrices: PVI, the JUE, CyUE, cJUE and scaled limits*, Nagoya Math. J. **174** (2004), 29–114, arXiv:math-ph/0204008.
- [26] S.N. Majumdar and G. Schehr, *Top eigenvalue of a random matrix: large deviations and third order phase transition*, J. Stat. Mech. Theory Exp. (2014), arXiv:1311.0580.
- [27] M.R. Douglas and V.A. Kazakov, *Large N phase transition in continuum QCD₂*, Phys. Lett. B **319** (1993), 219–230, arXiv:hep-th/9305047.
- [28] C. A. Tracy and H. Widom, *Level spacing distributions and the Airy kernel*, Comm. Math. Phys. **159** (1994), 151–174, arXiv:hep-th/9211141.
- [29] P. D. Draznev and E. B. Saff, *Constrained energy problems with applications to orthogonal polynomials of a discrete variable*, J. Anal. Math. **72** (1997), no. 1, 223–259.
- [30] A.B.J. Kuijlaars, *On the finite-gap ansatz in the continuum limit of the Toda lattice*, Duke Math. J. **104** (2000), no. 3, 433–462.
- [31] R. Koekoek, R. F. Swarttouw, and P. A. Lesky, *Hypergeometric orthogonal polynomials and their q -analogues*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2010.

INFN, SEZIONE DI FIRENZE, VIA G. SANSONE 1, SESTO FIORENTINO (FI), 50019, ITALY
E-mail address: colomo@fi.infn.it

STEKLOV MATHEMATICAL INSTITUTE, FONTANKA 27, ST. PETERSBURG, 191023, RUSSIA
E-mail address: agp@pdmi.ras.ru